Patch-Based Low-Rank Minimization for Image Denoising

Haijuan Hu, Jacques Froment, Quansheng Liu

***Abstract*—Patch-based sparse representation and low-rank approximation for image processing attract much attention in recent years. The minimization of the matrix rank coupled with the Frobenius norm data fidelity can be solved by the hard thresholding filter with principle component analysis (PCA) or singular value decomposition (SVD). Based on this idea, we propose a patch-based low-rank minimization method for image denoising, which learns compact dictionaries from similar patches with PCA or SVD, and applies simple hard thresholding filters to shrink the representation coefficients. Compared to recent patch-based sparse representation methods, experiments demonstrate that the proposed method is not only rather rapid, but also effective for a variety of natural images, especially for texture parts in images.**

***Index Terms*—Image denoising, patch-based method, low-rank minimization, principal component analysis, singular value decomposition, hard thresholding**

I. Introduction

IMAGE denoising is a classical image processing problem, but it still remains very active nowadays with the massive and easy production of digital images. We mention below some important works among the vast literature which deals with image denoising.

One category of denoising methods concerns transform-based methods, for example [1], [2]. The main idea is to calculate wavelet coefficients of images, shrink the coefficients and finally reconstruct images by inverse transform. These methods apply fixed transform dictionaries to whole images. However, fixed dictionaries do not generally represent whole images very well due to the complexity of natural images. Many image details are lost while being denoised.

Another category is related to patch-based methods first proposed in [3], which explores the non-local self-similarity of natural images. Inspired by this "patch-based" idea, the authors of K-SVD [4] and BM3D [5] proposed using dictionaries to represent small local patches instead of whole images so that sparsity of coefficients can be increased, where the dictionaries are fixed or adaptive, and compact or overcomplete. These methods greatly improve the traditional methods [1], [2], leading to very good performance. Since these works many similar methods have been proposed to improve the denoising process, such as LPG-PCA [6], ASVD [7], PLOW [8], SAIST [9], NCSR [10], GLIDE [11], and WNNM [12]. However, many proposed methods are computationally complex. For example, K-SVD uses overcomplete dictionaries for sparse representation, which is time-consuming. BM3D and LPG-PCA iterate the denoising process twice; SAIST and WNNM iterate about 10 times. The computational cost is directly proportional to the number of iterations.

At the same time, the low-rank matrix approximation has been widely studied and applied to image processing [13], [14], [15]. Many low-rank models have no explicit solution. However, the paper [13] proves that the nuclear norm minimization with the Frobenius norm data fidelity can be solved by a soft thresholding filter. (See also the paper [12] where an alternative proof is given.) Furthermore, with the help of Eckart-Young theorem [16], the paper [17] demonstrates that the solution of the exact low-rank matrix minimization problem (l_0 norm) can be obtained by a hard thresholding filter.

Inspired by the above theories, in this paper, a patch-based low-rank minimization (PLR) method is proposed for image denoising. First, similar patches are stacked together to construct similarity matrices. Then each similarity matrix is denoised by minimizing the matrix rank coupled with the Frobenius norm data fidelity. The minimizer can be obtained by a hard thresholding filter with principle component analysis (PCA) or singular value decomposion (SVD). The proposed method is rather rapid, since we use compact dictionaries which are more computationally efficient than over-completed dictionaries, and we do not iterate. Moreover, experiments show that the proposed method is as good as the state-of-the-art methods, such as K-SVD [4], BM3D [5], LPG-PCA [6], ASVD [7], PLOW [8], SAIST [9], and WNNM [12].

The rest of the paper is organized as follows. In Section II, we introduce our method. The experimental results are shown in Section III. Finally, this paper is concluded in Section IV.

II. Patch-Based Low-Rank Minimization

The noise model is:

$$\boldsymbol{v} = \boldsymbol{u} + \boldsymbol{\eta},$$

where $\boldsymbol{u}$ is the original image, $\boldsymbol{v}$ is the noisy one, and $\boldsymbol{\eta}$ is the Gaussian noise with mean 0 and standard deviation σ. The images $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\eta}$ are with size $M \times N$. Without loss of generality, we suppose that $M = N$.

A. Proposed Algorithm

Divide the noisy image $\boldsymbol{v}$ into overlapped patches of size $d \times d$. Denote the set of all these patches as $\boldsymbol{\mathcal{S}} = \{\boldsymbol{x}_i : i =$

Haijuan Hu is with Northeastern University at Qinhuangdao, School of Mathematics and Statistics, Hebei, 066004, China. She was with Univ Bretagne-Sud, CNRS UMR 6205, LMBA, Campus de Tohannic, F-56000 Vannes, France (e-mail: huhaijuan61@126.com).

Jacques Froment and Quansheng Liu are with Univ Bretagne-Sud, CNRS UMR 6205, LMBA, Campus de Tohannic, F-56000 Vannes, France (e-mail: Jacques.Froment@univ-ubs.fr; Quansheng.Liu@univ-ubs.fr).

$1, 2, \cdots, (N-d+1)^2\}$.

For each patch $\boldsymbol{x} \in \boldsymbol{\mathcal{S}}$, called reference patch, consider all the overlapped patches contained in its $n \times n$ neighborhood[1] (the total number of such patches is $(n-d+1)^2$ patches). Then choose the m ($m \geq d^2$) most similar patches (including the reference patch itself) to the reference patch among the $(n-d+1)^2$ patches. The similarity is determined by the l^2-norm distance.

Next, for each reference patch, its similar patches are reshaped as vectors, and stacked together to form a matrix of size $d^2 \times m$, called similarity matrix. The similarity matrix is denoted as $\boldsymbol{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_m)$, where columns of $\boldsymbol{S}$, i.e. $\boldsymbol{s}_i, i = 1, 2, \cdots, m$, are vectored similar patches. Then all the patches in the matrix $\boldsymbol{S}$ are denoised together using the hard thresholding method with the principal component (PC) basis, or equivalently, with the singular value decomposition (SVD) basis derived from the matrix $\boldsymbol{S}$; the detailed process will be given afterward. For convenience, we assume that the mean of the patches in $\boldsymbol{S}$, denoted by $\boldsymbol{s}_c := \frac{1}{m}\sum_{l=1}^{m} \boldsymbol{s}_l$, is 0. In practice, we subtract $\boldsymbol{s}_c$ from $\boldsymbol{s}_i$ to form the matrix $\boldsymbol{S}$, and add $\boldsymbol{s}_c$ to the final estimation $\bar{\boldsymbol{s}}_l$ of each patch.

Since the patches are overlapped, every pixel is finally estimated as the average of repeated estimates.

The process of denoising the matrix $\boldsymbol{S}$ is shown as follows. Firstly, we derive adaptive basis using PCA. The PC basis is the set of the eigenvectors of $\boldsymbol{S}\boldsymbol{S}^T$. Write the eigenvalue decomposition[2]

$$\boldsymbol{S}\boldsymbol{S}^T = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1} \tag{1}$$

with

$$\boldsymbol{P} = (\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_{d^2}), \boldsymbol{\Lambda} = \mathrm{diag}(m\lambda_1^2, m\lambda_2^2, \cdots, m\lambda_{d^2}^2),$$

where $\boldsymbol{g}_i$ denotes the i-th column of $\boldsymbol{P}$ and $\mathrm{diag}(c_1, c_2, \cdots)$ denotes the diagonal matrix with $(c_1, c_2, \cdots)$ on the diagonal. The PC basis is the set of the columns of $\boldsymbol{P}$, that is, $\{\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_{d^2}\}$.

The original patches $\boldsymbol{s}_i$ in the similarity matrix $\boldsymbol{S}$ are estimated as follows:

$$\bar{\boldsymbol{s}}_l = \sum_{k=1}^{d^2} a_k \langle \boldsymbol{s}_l, \boldsymbol{g}_k \rangle \boldsymbol{g}_k, \quad l = 1, 2, \cdots, m \tag{2}$$

where

$$a_k = \begin{cases} 1 & \text{if } \lambda_k^2 > t^2, \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

t being the threshold. Or equivalently, the matrix composed of estimated patches (2) can be written as

$$\bar{\boldsymbol{S}} := (\bar{\boldsymbol{s}}_1, \bar{\boldsymbol{s}}_2, \cdots, \bar{\boldsymbol{s}}_m) = \boldsymbol{P} h(\boldsymbol{\Lambda}) \boldsymbol{P}^{-1} \boldsymbol{S}, \tag{4}$$

with

$$h(\boldsymbol{\Lambda}) = \mathrm{diag}(a_1, a_2, \cdots, a_{d^2}). \tag{5}$$

Note that

$$\frac{1}{m}\sum_{l=1}^{m} (\langle \boldsymbol{s}_l, \boldsymbol{g}_k \rangle)^2 = \lambda_k^2 \tag{6}$$

after a simple calculation. Thus λ_k can be interpreted as the standard deviation of the basis coefficients.

We could also consider the singular value decomposition (SVD) of $\boldsymbol{S}$:

$$\boldsymbol{S} = \boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{Q}^T, \tag{7}$$

where $\boldsymbol{P}$ is chosen as the same orthogonal matrix in (1), $\boldsymbol{\Sigma}$ is a diagonal matrix, and $\boldsymbol{Q}$ (of size $m \times d^2$) has orthogonal columns such that $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$ with $\boldsymbol{I}$ the identity matrix. Then the denoised matrix (4) is equal to

$$\hat{\boldsymbol{S}} := \boldsymbol{P} H_{t\sqrt{m}}(\boldsymbol{\Sigma}) \boldsymbol{Q}^T, \tag{8}$$

where $H_{t\sqrt{m}}(\boldsymbol{\Sigma})$ is a diagonal matrix, with the diagonal of $H_{t\sqrt{m}}(\boldsymbol{\Sigma})$ obtained by the hard thresholding operator

$$H_{t\sqrt{m}}(\boldsymbol{\Sigma})_{kk} = \begin{cases} \boldsymbol{\Sigma}_{kk} & \text{if } \boldsymbol{\Sigma}_{kk} > t\sqrt{m}, \\ 0 & \text{otherwise,} \end{cases} \quad k = 1, 2, \cdots, d^2. \tag{9}$$

In fact, the equality of (4) and (8) can be demonstrated as follows. By the equations (1) and (7), we have $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^2$, and $\boldsymbol{P}^{-1}\boldsymbol{S} = \boldsymbol{\Sigma}\, \boldsymbol{Q}^T$. Furthermore, by the equations (5) and (9), we get $h(\boldsymbol{\Lambda})\boldsymbol{\Sigma} = H_{t\sqrt{m}}(\boldsymbol{\Sigma})$. Thus it follows that $\bar{\boldsymbol{S}} = \hat{\boldsymbol{S}}$.

B. *Low-Rank Minimization*

Theorem 2.1 stated below is an unconstrained version of the Eckart-Young theorem [16], and comes from Theorem 2(ii) in [17]. According to Theorem 2.1, it easily follows that

$$\hat{\boldsymbol{S}} = \arg\min_{X} \|\boldsymbol{S} - X\|_F^2 + mt^2 Rank(X), \tag{10}$$

where the minimum is taken over all the matrices X having the same size as $\boldsymbol{S}$, and $\|\cdot\|_F$ is the Frobenius norm. Hence the denoised matrix $\hat{\boldsymbol{S}}$ is the solution of the exact low-rank minimization problem.

Theorem 2.1: The following low-rank minimization problem

$$\hat{X} = \arg\min_{X} \|Y - X\|_F^2 + \mu\, Rank(X) \tag{11}$$

has the solution[3]

$$\hat{X} = U H_{\sqrt{\mu}}(\Sigma) V^T, \tag{12}$$

where $Y = U\Sigma V^T$ is the SVD of Y, and $H_{\sqrt{\mu}}$ is the hard thresholding operator

$$H_{\sqrt{\mu}}(\Sigma)_{kk} = \begin{cases} \Sigma_{kk} & \text{if } \Sigma_{kk} > \sqrt{\mu}, \\ 0 & \text{otherwise.} \end{cases}$$

C. *Choice of the Threshold*

The choice of the threshold t in (3) is crucial for the proposed algorithm. We study it by minimizing the mean squared error of estimated values of vectored patches $\boldsymbol{s}_l, (l = 1, 2, \cdots, m)$ in a similarity matrix $\boldsymbol{S}$. Denote

$$\boldsymbol{s}_l = \boldsymbol{u}_l + \boldsymbol{\eta}_l,$$

where $\boldsymbol{u}_l$ and $\boldsymbol{\eta}_l$ are the vectored patches of the true image and the noise corresponding to $\boldsymbol{s}_l$ respectively.

[1] The reference patch is located at the center of the neighborhood, if the parities of d and n are the same; otherwise, the reference patch is located as near as possible to the center of the neighborhood.

[2] We assume that the matrix $\boldsymbol{S}\boldsymbol{S}^T$ has full rank, and it has no identical eigenvalues, which are generally true in practice.

[3] Strictly speaking, if none of the singular values of Y equals with $\sqrt{\mu}$, the solution is unique, which is generally true in practice.

By (2) or (4), it can be easily obtained that

$$\|\bar{\boldsymbol{s}}_l - \boldsymbol{u}_l\|^2 = \sum_{k=1}^{d^2} (a_k - 1)^2 (\langle \boldsymbol{g}_k, \boldsymbol{u}_l \rangle)^2 + \sum_{k=1}^{d^2} a_k^2 (\langle \boldsymbol{g}_k, \boldsymbol{\eta}_l \rangle)^2. \quad (13)$$

Assume that the PC basis $\{\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_{d^2}\}$ only depends on the true value vectors $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_m\}$ and hence is independent of $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_m\}$. Then

$$\mathbb{E}(\langle \boldsymbol{g}_k, \boldsymbol{\eta}_l \rangle)^2 = \sigma^2. \quad (14)$$

Let

$$\theta_k^2 = \frac{1}{m} \sum_{l=1}^{m} (\langle \boldsymbol{g}_k, \boldsymbol{u}_l \rangle)^2. \quad (15)$$

Then by (6), we obtain

$$\mathbb{E}(\lambda_k^2) = \theta_k^2 + \sigma^2. \quad (16)$$

Thus, from (13), (14), and (15), it follows that

$$\frac{1}{m} \sum_{l=1}^{m} \|\bar{\boldsymbol{s}}_l - \boldsymbol{u}_l\|^2 \approx \sum_{k=1}^{d^2} (a_k - 1)^2 \theta_k^2 + \sigma^2 \sum_{k=1}^{d^2} a_k^2. \quad (17)$$

After a simple calculation, the optimal value for a_k is

$$\hat{a}_k = \begin{cases} 1 & \text{if } \theta_k^2 > \sigma^2, \\ 0 & \text{otherwise.} \end{cases}$$

Since $\lambda_k^2 \approx \theta_k^2 + \sigma^2$ by (16), the optimal value of the threshold in (3) is $t^2 \approx 2\sigma^2$. In practice, we find that $t = 1.5\sigma$ is a good choice.

III. Experimental Results

In this section, we compare the performance of our PLR method with those of state-of-the-art methods, including the highly competitive method WNNM [12] proposed very recently. Standard gray images are utilized to test the performance of methods. For the simulation, the level of noise is supposed to be known, otherwise there are methods to estimate it; see e.g. [18]. For each image and each level of noise, all the methods are applied to the same noisy images.

For our algorithm, the patch size is set to $d = 7$, the size of neighborhoods for selecting similar patches is set to $n = 35$, and the number of similar patches in a similarity matrix is chosen as $m = 5d^2$. Image boundaries are handled by assuming symmetric boundary conditions. For the sake of computational efficiency, the moving step from one reference patch to its neighbors both horizontally and vertically is chosen as the size of patches, that is, 7. For other comparison algorithms, we utilize the original codes released by theirs authors.

In Table I, we compare the PSNR (Peak Signal-to-Noise Ratio) values of our PLR method with other methods. The PSNR value is defined by

$$\text{PSNR } (\bar{\boldsymbol{v}}) = 20 \log_{10} \frac{255N}{\|\bar{\boldsymbol{v}} - \boldsymbol{u}\|_F} \text{dB},$$

where $\boldsymbol{u}$ is the original image, and $\bar{\boldsymbol{v}}$ the restored one. As can be seen in Table I, our method is generally better than K-SVD [4], LPG-PCA [6] and PLOW [8], and sometimes even better than BM3D [5]. Furthermore, for the visual comparisons, our method is also good. For example, as can be seen in Fig.1, our method preserves the texture parts in Lena and Barbara the best among all the methods.

TABLE I
PSNR values for removing noise for our PLR and other methods. Cam is the Cameramen image

Image	Lena	Barbara	Peppers	Boats	Bridge	House	Cam
$\sigma = 10$							
K-SVD[4]	35.50	34.82	34.23	33.62	30.91	35.96	33.74
LPGPCA[6]	35.72	35.46	34.05	33.61	30.86	36.16	33.69
ASVD[7]	35.58	35.58	33.55	33.26	27.76	36.46	31.62
PLOW[8]	35.29	34.52	33.56	32.94	29.88	36.22	33.15
PLR	35.90	35.50	34.28	33.76	30.78	36.57	33.73
BM3D[5]	35.90	35.39	34.68	33.88	31.06	36.71	34.18
SAIST[9]	35.87	35.69	34.76	33.87	31.03	36.52	34.28
WNNM[12]	36.02	35.92	34.94	34.05	31.16	36.94	34.44
$\sigma = 20$							
K-SVD[4]	32.38	31.12	30.78	30.37	27.03	33.07	30.01
LPGPCA[6]	32.61	31.69	30.50	30.26	26.84	33.10	29.77
ASVD[7]	33.21	32.96	30.56	31.79	25.51	33.53	29.33
PLOW[8]	32.70	31.48	30.52	30.36	26.56	33.56	29.59
PLR	33.03	32.12	30.90	30.64	27.20	33.36	30.12
BM3D[5]	33.03	32.07	31.28	30.85	27.14	33.77	30.48
SAIST[9]	33.07	32.43	31.28	30.78	27.20	33.80	30.40
WNNM[12]	33.10	32.49	31.53	30.98	27.29	34.01	30.75

TABLE II
Running time in second for our PLR and other methods to remove noise with images of size 256×256

K-SVD	LPG-PCA	ASVD	PLOW	**PLR**	SAIST	WNNM
210	138	337	43	2	25	134

To have a clear comparison of complexities of different methods, we compare the average CPU time to remove noise with $\sigma = 20$ for the testing images of size 256×256: Peppers, House and Cameraman. All the codes are written in M-files and run in the platform of MATLAB R2011a on a 3.40GHz Intel Core i7 CPU processor. We do not include BM3D for comparison since the original code of BM3D contains MEX-files. The running time is displayed in second in Table II. The comparisons clearly show that the proposed method is much faster than the others.

IV. Conclusion

In this paper, a patch-based low-rank minimization method for image denoising is proposed, which stacks similar patches into similarity matrices, and denoises each similarity matrix by seeking the minimizer of the matrix rank coupled with the Frobenius norm data fidelity. The minimizer can be obtained by a hard threshoding filter with principle component basis or left singular vectors. The proposed method is not only rapid, but also effective compared to recently reported methods.

References

[1] D. Donoho and J. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[2] R. Coifman and D. Donoho, “Translation-invariant denoising,” *Wavelets and statistics*, vol. 103, pp. 125–150, 1995.

[3] A. Buades, B. Coll, and J. Morel, “A review of image denoising algorithms, with a new one,” *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[4] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

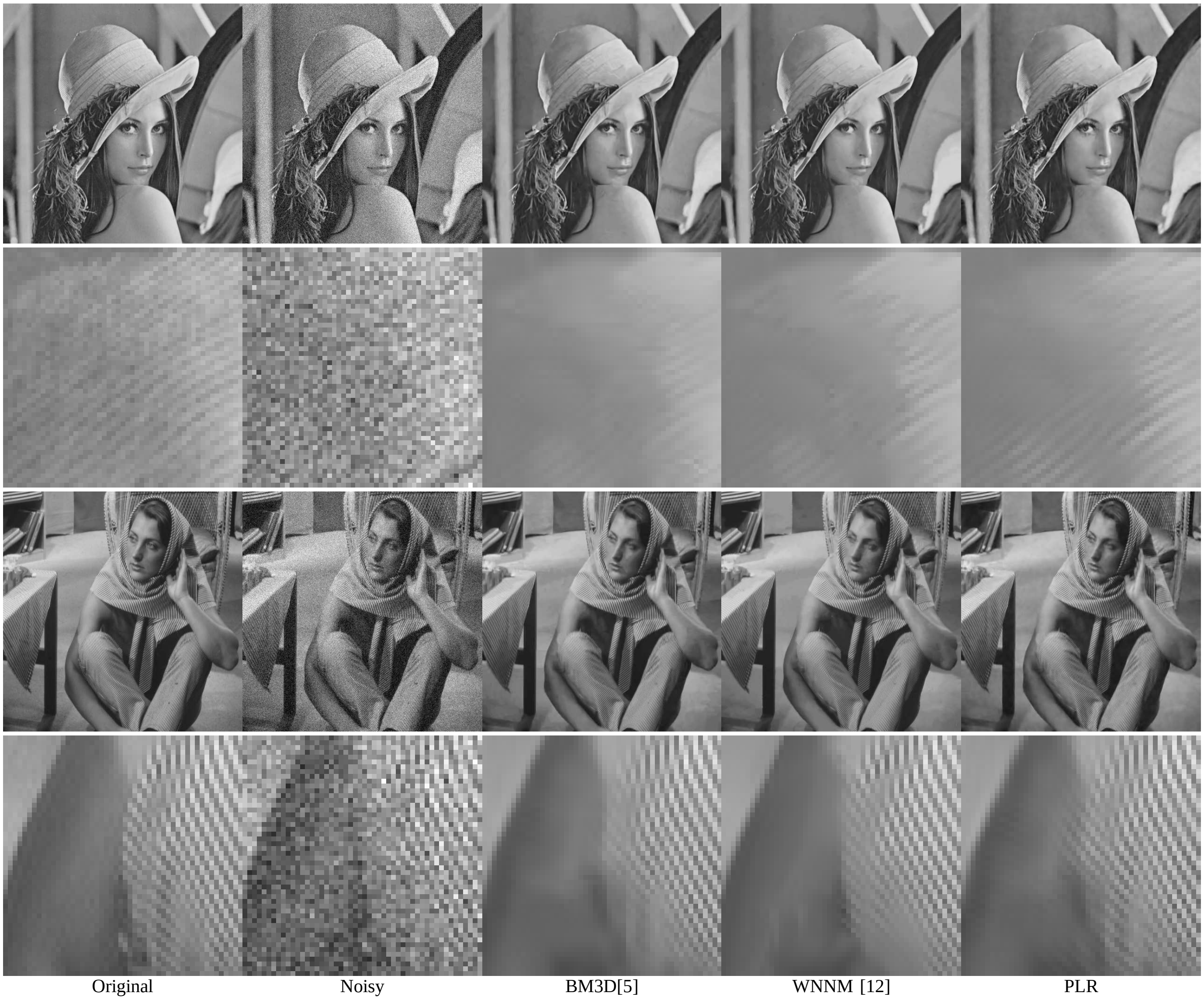

Fig. 1. Compare denoised images Lena and Barbara by our method and other methods for $\sigma = 20$. From left to right, the images are original images, noisy images, images denoised by BM3D, WNNM, and our PLR method. To make the differences clearer, the second row and the bottom row display parts of Lena images and Barbara images extracted from the first row and third row respectively.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[6] L. Zhang, W. Dong, D. Zhang, and G. Shi, "Two-stage image denoising by principal component analysis with local pixel grouping," *Pattern Recognition*, vol. 43, no. 4, pp. 1531–1549, 2010.

[7] Y. He, T. Gan, W. Chen, and H. Wang, "Adaptive denoising by singular value decomposition," *IEEE Signal Process. Lett.*, vol. 18, no. 4, pp. 215–218, 2011.

[8] P. Chatterjee and P. Milanfar, "Patch-based near-optimal image denoising," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1635–1649, 2012.

[9] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation: a low-rank approach," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 700–711, 2013.

[10] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration." *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1620–1630, 2013.

[11] H. Talebi and P. Milanfar, "Global image denoising," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 755–768, 2014.

[12] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014.

[13] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optimiz.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[14] H. Schaeffer and S. Osher, "A low patch-rank interpretation of texture," *SIAM J. Imaging Sci.*, vol. 6, no. 1, pp. 226–262, 2013.

[15] Y. Peng, J. Suo, Q. Dai, and W. Xu, "Reweighted low-rank matrix recovery and its application in image restoration," *IEEE Trans. Cybern.*, vol. 14, no. 12, pp. 2418–2430, 2014.

[16] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[17] J.-B. Hiriart-Urruty and H. Y. Le, "From eckart and young approximation to moreau envelopes and vice versa," *RAIRO-Operations Research*, vol. 47, no. 03, pp. 299–310, 2013.

[18] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *J. Roy. Stat. Soc. B*, vol. 59, no. 2, pp. 319–351, 1997.